

Domain Modeling for Machine Translation

J. Joachim Quantz¹, Manfred Gehrke², Uwe Küssner¹

¹TU Berlin, Project KIT-VM11, {jjq,uk}@cs.tu-berlin.de

²Siemens AG, Manfred.Gehrke@zfe.siemens.de

1 Introduction

In this paper we present the domain model used in the VERBMOBIL [10] project, a project concerned with face-to-face dialogue interpreting funded by the German Ministry of Research and Technology. The current version of the domain model, which is used in the VERBMOBIL demonstrator, has been implemented in the Description Logic (DL) system FLEX¹ and contains approximately 300 concepts and 170 roles [16]. Instead of describing the details of this model we will rather motivate its main design criteria and illustrate its functionality by considering its use in the disambiguation process.

In the next section we sketch the task of domain modeling in general, describe the particular requirements arising in the context of Machine Translation, and discuss the relationship between domain modeling and lexical semantics. In Section 3 we show how a contrastive analysis of source and target language forms the basis for a transfer-oriented conceptual hierarchy. Section 4 then illustrates how conceptual information is used to preferentially order the output of the speech parser.

2 The Task of Domain Modeling

Domain models are used in Natural Language Processing systems to provide domain-specific background knowledge. This is usually achieved by defining the concepts occurring in a particular application, i.e. by modeling their relevant properties and the hierarchical relations between them. Obviously, the main problem consists in deciding which aspects are relevant for a particular application. In [16] we claim that a distinction should be made between the *representation* task and the *interpretation* task².

Given this distinction, we think that in Machine Translation in general, and in the VERBMOBIL demonstrator in particular, semantics is mainly needed to support the interpretation task. Though semantic representation is used as an input for Transfer and Generation, there is no need to use it immediately for drawing inferences. Contrast this with Information Systems, in which the semantic content of a query has to be evaluated wrt the stored information; or with text understanding systems

¹ FLEX [17] is an extension of the DL system BACK [7] developed in the VERBMOBIL project.

² Roughly speaking, the representation task consists in choosing a representation which captures the relevant information aspects of an expression, while the interpretation task consists in automatically deriving such a representation (see also [1, p. 9]).

in which the information contained in a text has to be represented semantically in order to answer queries.

As a consequence, we opt for an approach more in the spirit of *AI semantics* as opposed to *LL Semantics*. The distinction between AI semantics and LL semantics has been suggested in [13]. AI semantics are based on representation formats as *Semantic Nets*, *Frames*, or *Scripts*. LL semantics are based on Montague's initial work and comprise *Montague Semantics*, *Discourse Representation Theory*, and *Situation Semantics*. Pinkal sketches some of the main differences between both research areas concerning the subject of research, the ontological and methodological assumptions, etc. We think that the most important difference between the two research paradigms concerns the *status of semantic representations*. Not surprisingly, AI semantics is interested in the computational aspects of semantic representation, more precisely it addresses two questions:

1. Given an NL expression, how can the computer determine an appropriate semantic representation?
2. Given the semantic representation of an NL expression, how can the computer draw inferences from it?

LL semantics, on the other hand, is more interested in the logical aspects of semantic representations, e.g. in the problem of formally specifying the appropriate semantic representations of NL expressions. Appropriateness is here usually understood as correctly capturing the truth conditions of an expression.

Given the dominance of truth-conditional semantics in the last decades, it is not surprising that many approaches to lexical meaning rely on the truth-condition-testing method. Consider, for example, the definition of cognitive synonymy in [4, p. 88]:

X is a cognitive synonym of Y, if (i) X and Y are syntactically identical, and (ii) any grammatical declarative sentence S containing X has equivalent truth-conditions to another sentence S', which is identical to S except that X is replaced by Y.

Though we think that this is a useful test criterion for synonymy it is not a real definition since it reduces synonymy of lexemes to equivalence of truth conditions, i.e. to analyticity (see [18, p. 28ff]). In fact recourse to truth conditions does not help us much in modeling conceptual contents, because if we are uncertain about the exact meaning of an expression we will not be able to specify the exact truth conditions of a sentence containing it. Furthermore, truth conditions are usually underdetermined by sentences, i.e. they are only determined given a particular choice of the conceptual contents of the expressions occurring in the sentences (see, for example, [3, p. 251]) or [12, p. 12ff].

One of the main problems is therefore to decide *how many* conceptual contents we have to assign to an expression? Consider an example taken from the literature, namely the meaning of the lexeme 'open' discussed in [19, p. 145]:

- (1) a. Tom opened the door.
b. Sally opened her eyes.
The carpenters opened the wall.

- d. Sam opened his book on page 37.
- e. The surgeon opened the wound.

Searle claims that

... the word “open” has the same literal meaning in all five of these occurrences. Anyone who denied this would soon be forced to hold the view that the word “open” is indefinitely or perhaps even infinitely ambiguous since we can continue the example; and indefinite ambiguity seems an absurd result. [19, p. 146]

He gives additional examples in which one might argue that the lexeme ‘open’ has a different meaning:

- (2) a. The chairman opened the meeting.
- b. The artillery opened fire.
- c. Bill opened a restaurant,

Searle uses these examples to show that understanding requires a *preintentional Background* and that the literal meaning of a sentence is thus not a context-free notion, but is relative to a set of preintentional Background assumptions and practices [19, p. 145f].

Let us summarize the discussion:

- 1. the conceptual content of expressions is context dependent;
- 2. it is problematic to decide how many concepts to assign to an expression (truth-condition tests are helpful but do not resolve all problematic cases).

Since we are not primarily interested in truth conditions anyway, we propose to base the assignment of concepts to expressions on the requirements of our particular application, namely translation from German into English.

3 Transfer-Oriented Conceptual Hierarchies

A straightforward criterion for deciding how many conceptual contents to assign to a source-language expression is to provide a separate concept for each possible target-language translation. Consider the examples (1) and (2)—the German translation for ‘open’ in (1) is ‘öffnen’, whereas in (2) it is ‘eröffnen’. This would be a reason to distinguish the conceptual content of ‘open’ expressed in (1) and (2). In general, we would thus have *m-to-n* mappings between NL expressions and concepts.³

Note that the structure of the conceptual hierarchy to be modeled thus depends on the particular source and target languages. Based on *contrastive analyses* it can be decided whether a concept has to be included into the domain model or not. Note further, that it is in principle possible to extend such a bilingually motivated hierarchy by adding new source and target languages, i.e. by adding mappings and integrating the additionally required conceptual distinctions [6, 8].

³ We thus assume that the so-called translational mismatches [9] are not the exception but rather the average case.

Another source of requirements for conceptual distinction arises from the general problem of conceptual disambiguation. Since inferences involving complex background knowledge are usually not available, conceptual disambiguation has to be based on heuristics, i.e. partial information and incomplete reasoning. The most popular way of doing conceptual disambiguation is by means of *selectional restrictions*. Roughly speaking, conceptual disambiguation is performed in this approach by considering the arguments of an expression, or vice versa the functor taking an expression as argument. Note that this is a rather limited way of taking context into account, but it is exactly this limitation which makes it computationally feasible.

For illustration consider again the examples (1) and (2). To distinguish between the two readings of ‘open’ we can check whether the argument of ‘open’ is a thing or an event. If it is a thing we map ‘open’ to the concept expressed by ‘öffnen’, if it is an event we map it to the concept expressed by ‘eröffnen’.

We will illustrate this by considering the translation of prepositions occurring in free adjuncts. First, a contrastive analysis yields for each German preposition a set of corresponding English prepositions. Having established the set of possible translations of a German preposition, the next step is to determine the information relevant for choosing the appropriate translation in a given utterance. For prepositions occurring in free adjuncts the relevant information is given by

1. the *internal argument*, i.e. the noun phrase constituting the prepositional phrase together with the preposition;
2. the *external argument*, i.e. the phrase modified by the prepositional phrase.

Both syntactic and semantic information about the internal and the external argument can be relevant for translation.

Having sketched this general methodology, it should be noted that there are two main problems concerning the details. For one thing, there are conceptual differences which are irrelevant for translation from German to English but seem too important to be ignored in the conceptual hierarchy. Thus the German preposition ‘in’ can be used with spatial and temporal noun phrases and is in both cases translated by the English preposition ‘in’. This indicates that focussing solely on the linguistic level, i.e. transfer from German to English can produce “conceptually weird” hierarchies.

On the other hand, contrastive analyses might yield “overspecified” concepts. Consider the translation pair

- (3) a. Mein Büro ist *im* zweiten Stock.
- b. My office is *on* the second floor.

It could be argued that ‘Stock’ in German is conceptualized as a three-dimensional container, whereas ‘floor’ in English is conceptualized as a two-dimensional plane. Instead of modeling these language-specific distinctions in the conceptual hierarchy it seems more appropriate to leave the selection of the preposition ‘on’ to the generation component.

To summarize this section, the conceptual disambiguation of an expression can thus be triggered by two different requirements:

1. in order to determine the corresponding target-language expression (representation task);

2. in order to determine the conceptual content of *another expression*, e.g. a functor or an argument of the expression (interpretation task).

In the following section we describe a particular interpretation task in more detail, namely the preferential ordering of alternative parses of an expression.

4 Conceptual Support for Speech Parsers

While semantic evaluation and transfer use quite specific information from the domain model, the detection of sortal mismatches on the parsing results relies on more general conceptual distinctions. Examining the word-hypotheses lattices produced by a recognizer besides perfect reproductions of the spoken utterances one finds

1. syntactically and semantically correct sentences, which do not match the spoken utterance; detecting such cases would require a precise pragmatic analysis, thus one just can hope that such cases somehow fit into the ongoing discourse;
2. syntactically correct, but semantically deviant sentences, which can be detected by the application of selectional restrictions as shown below;
3. strings of words with no recognizable syntactic structure, which are discarded by the grammar.

Though the application of selectional restrictions may lead to wrong results in cases of type coercion or metonymies, they are nevertheless necessary to detect recognizer results of the second type. Consider the following examples where the first one is a recognizer mismatch, while the second is a case of type coercion (violations of sortal restrictions are emphasized):

- (4) Dann muß ich *drei Uhr* auf den Freitag verschieben.
I have to postpone *three o'clock* to friday.
- (5) Dann muß ich *Dreyer* auf den Freitag verschieben.
I have to postpone *Dreyer* to friday.
lit.: I have to postpone the meeting with Dreyer to friday.

If the process of semantic construction should not be burdened with the task to select the semantically sensible readings out of the parsing results, a solution to this dilemma, i.e. discarding meaningless utterances, but still have a chance to interpret cases of type coercion, is to apply the test of the selectional conditions as a soft constraint, which gives each selectional clash a penalty instead of immediately rejecting it. Such a strategy also contributes to the robustness of the system. Thus the application of selectional restrictions results in an ordering of the results of the parser, where the “sortally” best reading is presented first to the process of semantic construction. To provide tests for selectional restrictions the grammar rules have been augmented by two kinds of features. At first the selectional conditions, which finally result in subsumption tests between concepts⁴ of the domain model, have to be percolated from the lexicon to the respective constituents. Secondly, whenever a daughter

⁴ Wrt the *m-to-n*-mapping between lexemes and concepts exemplified in Section 3 the sort of a lexeme is often enough a rather general concept, located near the top of the concept hierarchy. In the course of conceptual disambiguation it will be refined to a more specific concept depending on the context of actual usage as e.g. described in [2, 14, 15].

which serves as an argument or adjunct, is adjoined to a constituent, it has to be checked whether the selectional conditions are met.

Lexical entries with an argument structure, such as verbs, adjectives, and prepositions also state conditions on the sorts of their arguments. There are two types of tests to check selectional restrictions:

- A subsumption test between the sort of an argument position and the sort of the possible argument, where the first one has to subsume the second one. The same holds between the sort of a phrase and the sort of an adjunct.
- Checking whether a specific relation (a FLEX-role) holds between two arguments, as it is the case for semantically empty verbs such as *to be* or *to have*. A distinctive feature for *to be* in the definitorial reading is that the sort of the subject has to be subsumed by the sort of predicate noun.

Since the sorts are FLEX-concepts and the subsumption test is performed using the subsumption process of FLEX, which differs from the notion of subsumption in unification, the test of the selectional restrictions is performed after the application of the parsing process. It might be argued that the typed feature structures (TFS) as being used in HPSG or the mapping of ISA-Hierarchies into unifiable term structure as proposed in [11] allows a kind of subsumption inside the unification process. But these representation structures are less expressive than a description logic (DL) like FLEX, so that the mapping from a DL-structure into a TFS to avoid the construction and maintenance of multiple models, which provide the domain knowledge, will throw away some necessary information.

The application of the selectional tests after the parser process has the drawback that the analysis is not directly cancelled, when a sortal mismatch is detected. But this strategy opens the opportunity to trigger further reasoning processes, as e.g. for interpreting type coercion on sortally illegal analyses, if no correct analysis has been found. It should also be noted that the proper analysis of type coercion effects may depend on the right context of the respective utterance. The chosen strategy has also the advantage that the application of selectional tests can take into account global information from the utterance context, while the immediate application of selectional restriction is restricted to local information.

As mentioned above the selectional conditions are soft constraints modeled as a bonus system. Each successful application of such a condition increments the bonus counter. The bonus system is additionally refined by giving preferred constructions an extra bonus (Cf. [5]). Normalizing the total bonus figure with the number of all applications of selectional conditions of the utterance gives a measure for the selectional quality of an analysis. Thus we get the following "soft" quality criteria:

$$\frac{N(\text{successes})}{N(\text{tests})} = \begin{cases} f \geq 1 & \text{literal meaning} \\ < 1 & \text{non-literal meaning or acoustic recognition error} \end{cases}$$

Additionally the bonus system is used to get a better treatment of pronouns. With the exception of 1st and 2nd person personal pronouns and some other pronouns, which refer explicitly to persons, the sort of a pronoun ought to be subsumable by any other sortal concept. The only concept that fulfills this condition is the concept 'nothing', which is very undesirable, because it also denotes a mismatch. Thus the remaining pronouns get a special signature onto which the subsumption test is not

applied and their bonus is a figure slightly smaller than one. The reason for this smaller figure is twofold. On the one hand the sortal appropriateness of this analysis has still to be confirmed by the reference identification process of the DRT. On the other hand, if the grammar allows a pronoun being analyzed as a NP, as in

- (6) a. an den (termin) habe ich nicht gedacht,
- b. I forgot that date.

and when the definite article and the pronouns have the same form, then there are cases where it is by syntactic means nearly impossible not to analyse a "... Pron N ..."-sequence also as a sequence of two NPs. By the smaller figure for pronouns the interpretation of such a sequence as "... Art Noun ..." gets a better score, which is in most cases the appropriate reading.

Given the Verbmobil test corpus with approximately 200 turns, i.e. about 400 sentences, as evaluation basis, the test of selectional restrictions increases the total parsing time by 5 %. For a part of this corpus (76 utterances) that syntactic reading of each utterance has been determined which is the intended one in the respective dialogue. An application of the parser without testing selectional restrictions on this smaller corpus with written input gives the intended reading as the first reading in 49 (64 %) of all cases, while with selectional restrictions the first reading is the intended one in 71 (95 %) cases. Most of the remaining cases, where the correct reading is not the first one, are due to either PP-attachment (adjunct or modifier) or the lack of prosodic information, where the boundary between two sentences of an utterance is misplaced in the first reading. The evaluation on word lattices gives similar results. But one has to consider, that sortally deviant utterances are not discarded, they just have a worse scoring (< 1). When there is just one result later stages of evaluation have to decide whether to do some repair due to acoustic mismatches or to interpret type coercion.

5 Conclusion

We have shown how a domain model can provide relevant background knowledge in a Machine Translation system and can thus be used to support both the representation and the interpretation task. The main problem in designing such a domain model is to decide which concepts and properties are relevant in the particular application. We have argued that this decision should be based on requirements arising both in the representation task and the interpretation task. We have shown how a contrastive analysis of source language and target language forms the basis for a transfer-oriented conceptual hierarchy. This hierarchy can then be refined by taking into account requirements from the interpretation task. We have illustrated such requirements by presenting the preferential ordering of the speech-parser output in the VERBMOBIL project.

Acknowledgements

This work was funded/partially funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the Verbmobil Project

under Grants 01 IV 101 Q 8 and 01 IV 102 A 0. The responsibility for the contents of this study lies with the authors.

References

- [1] J. Allen, *Natural Language Understanding*, Menlo Park: Benjamin/Cummings, 1987
- [2] M. Bierwisch, "Semantische und konzeptuelle Repräsentation lexikalischer Einheiten", *Studia grammatica XII*, 61-99
- [3] P. Bosch, "Lexical Meaning Contextualized", in G. Hoppenbrouwers, A. Weijters, P. Seuren (eds), *Meaning and the Lexicon*, Dordrecht: Foris, 1985, 251-258
- [4] D.A. Cruse, *Lexical Semantics*, Cambridge: Cambridge University Press, 1986
- [5] M. Gehrke, *Sorting Syntactic Analysis by Selectional Restrictions*, Verbmobil Report (in preparation), Siemens AG, 1995
- [6] R. Henschel, J. Bateman, "The Merged Upper Model: A Linguistic Ontology for German and English", *COLING-94*, 803-809
- [7] T. Hoppe, C. Kindermann, J.J. Quantz, A. Schmiedel, M. Frischer, *BACK V5 Tutorial & Manual*, KIT Report 100, Technische Universität Berlin, 1993
- [8] E. Hovy, S. Nirenburg, "Approximating an Interlingua in a Principled Way", in *Proceedings of the, DARPA Speech and Natural Language Workshop*, Hawthorne, NY, Feb. 1992
- [9] M. Kameyama, R. Ochitani, S. Peters, "Resolving Translation Mismatches With Information Flow", *ACL'91*, 193-200
- [10] M. Kay, J.M. Gawron, P. Norvig, *Verbmobil: A Translation System for Face-to-Face Dialog*, CSLI Lecture Notes 33 August 1991
- [11] C. Mellish, "Implementing Systemic Classification by Unification", *Computational Linguistics* **14**, 40-51, 1988
- [12] M. Pinkai, *Logik und Lexikon*, Berlin: de Gruyter, 1985
- [13] M. Pinkai, *Neuere Semantikmodelle für die Verarbeitung natürlicher Sprache*, Universität Hamburg: FBI-HH-M-164/88, Dezember 1988
- [14] J. Pustejovsky, "Current Issues in Computational Lexical Semantics", *EACL-89*, xvii-xxv
- [15] J. Pustejovsky, "Type Coercion and Lexical Selection", in J. Pustejovsky (Ed.), *Semantics and the Lexicon*, Dordrecht: Kluwer, 1993, 73-94
- [16] J.J. Quantz, M. Gehrke, U. Küssner, B. Schmitz, *The VERBMOBIL Domain Model*, KIT Report 122, VERBMOBIL Report 29, Technische Universität Berlin, 1994
- [17] J.J. Quantz, G. Dunker, F. Bergmann, I. Kellner, *The FLEX System*, KIT Report in preparation, Technische Universität Berlin, 1995
- [18] W. v.O. Quine, "Two Dogmas of Empiricism", in W. v.O. Quine, *From a Logical Point of View*, Cambridge (Mass.): Harvard University Press, 1980²
- [19] J.R. Searle, *Intentionality*, Cambridge: Cambridge University Press, 1983